Journal of the American Heart Association

ORIGINAL RESEARCH

Modeling *SMAD2* Mutations in Induced Pluripotent Stem Cells Provides Insights Into Cardiovascular Disease Pathogenesis

Tarsha Ward, PhD; Sarah U. Morton, MD, PhD; Gabriela Venturini, PhD; Warren Tai, MD; Min Young Jang, MD; Joshua Gorham; Dan Delaughter, PhD; Lauren K. Wasson, PhD; Zahra Khazal; Jason Homsy, MD, PhD; Bruce D. Gelb, MD; Wendy K. Chung, MD, PhD; Benoit G. Bruneau, PhD; Martina Brueckner, MD; Martin Tristani-Firouzi, MD; Steven R. DePalma, PhD; Christine Seidman, MD* J. G. Seidman, PhD*

BACKGROUND: SMAD2 is a coregulator that binds a variety of transcription factors in human development. Heterozygous *SMAD2* loss-of-function and missense variants are identified in patients with congenital heart disease (CHD) or arterial aneurysms. Mechanisms that cause distinct cardiovascular phenotypes remain unknown. We aimed to define transcriptional and epigenetic effects of *SMAD2* variants and their role in CHD. We also assessed the function of *SMAD2* missense variants of uncertain significance.

METHODS AND RESULTS: Rare *SMAD2* variants (minor allele frequency $\leq 10^{-5}$) were identified in exome sequencing of 11 336 participants with CHD. We constructed isogenic induced pluripotent stem cells with heterozygous or homozygous loss-of-function and missense *SMAD2* variants identified in CHD probands. Wild-type and mutant induced pluripotent stem cells were analyzed using bulk RNA sequencing, chromatin accessibility (Assay for Transposase-Accessible Chromatin With Sequencing), and integrated with published SMAD2/3 chromatin immunoprecipitation data. Cardiomyocyte differentiation and contractility were evaluated. Thirty participants with CHD had heterozygous loss-of-function or missense *SMAD2* variants. *SMAD2* haploinsufficiency altered chromatin accessibility at promoters and dysregulated expression of 385 SMAD regulated genes, including 10 CHD-associated genes. Motifs enriched in differential Assay for Transposase-Accessible Chromatin peaks predicted that *SMAD2* haploinsufficiency disrupts interactions with transcription factors NANOG (homeobox protein NANOG), ETS, TEAD3/4 (transcriptional enhanced associate domain 3/4), CREB1 (cAMP response element binding protein 1), and AP1 (activator protein 1). Compared with *SMAD2*-haploinsufficient cells, induced pluripotent stem cells with R114C or W274C variants exhibited distinct and shared chromatin accessibility and transcription factor binding changes.

CONCLUSIONS: *SMAD2* haploinsufficiency disrupts transcription factor binding and chromatin interactions critical for cardiovascular development. Differences between the molecular consequences of loss-of-function and missense variants likely contribute to phenotypic heterogeneity. These findings indicate opportunities for molecular analyses to improve reclassification of *SMAD2* variants of uncertain clinical significance.

Key Words: congenital heart disease ■ CRISPR/Cas9 gene editing ■ induced pluripotent stem cells ■ mechanisms of transcriptional regulation ■ SMAD2

Correspondence to: J. G. Seidman, PhD, Harvard Medical School, New Research Building Room 256, 77 Avenue Louis Pasteur, Boston, MA 02115. Email: seidman@genetics.med.harvard.edu

*C. Seidman and J. G. Seidman contributed equally.

This article was sent to June-Wha Rhee, MD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.124.036860

For Sources of Funding and Disclosures, see page 12.

© 2025 The Author(s). Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

JAHA is available at: www.ahajournals.org/journal/jaha

RESEARCH PERSPECTIVE

What Is New?

- Exome sequence analysis of ~11 000 patients with congenital heart disease identified 31 subjects with de novo variants or rare inherited variants predicted to disrupt *SMAD2* transcriptional regulation may contribute to cardiac malformations and vascular anomalies.
- *SMAD2* haploinsufficiency in induced pluripotent stem cells impairs cardiac contractility and alters epigenetic and transcriptional processes that cause dysregulation of genes required for normal heart development; moreover, we have made isogenic induced pluripotent stem cells carrying 2 different rare *SMAD2* missense variants of unknown significance (W274C and R114C). Both disrupt SMAD2 function, suggesting that these variants can cause congenital heart disease.

What Question Should Be Addressed Next?

- These data suggest that, similar to *SMAD2* haploinsufficiency, heterozygous missense variants, W274C and R114C, disrupt *SMAD2*-mediated mechanisms that contribute to congenital heart disease phenotypes. However, the functional impact of 23 additional heterozygous *SMAD2* missense variants identified through exome sequence analysis of congenital heart disease probands remains unknown.
- Are there additional cellular and molecular assays that can be developed to improve the efficiency of reclassifying variants of uncertain significance?

Nonstandard Abbreviations and Acronyms

iPSC	induced pluripotent stem cells
iPSC-CMs	induced pluripotent stem cell-derived cardiomyocyte
LoF	loss of function
PGP1	personal genome project 1

Congenital heart disease (CHD), the most common human congenital anomaly, arises from maldevelopment of the embryonic heart.[1–4] Exome and genome sequencing of CHD probands and trios (proband and unaffected parents) demonstrate that ~40% of cases carry likely pathogenic variants in 1 of ~200 CHD genes,[4–7] thereby implicating critical molecules involved in lineage commitment, differentiation, and maturation of the heart. These observations also imply questions about the developmental impact by rare damaging missense variants that may also evoke malformations in patients with unexplained CHD.

Ongoing studies of loss-of-function (LoF) variants in ~130 CHD genes that result in haploinsufficiency allow classifications as pathogenic. However, missense variants in these genes that are present at low frequencies (minor allele frequency $\leq 10^{-5}$) or absent from population-based sequence databases[4–6,8] are variants of uncertain significance due to a lack of understanding the biological impact of amino acid substitutions. Reproducible functional assays in cell or animal models that distinguish damaging from benign missense variants could substantially enhance reclassification and advance insights into CHD mechanisms. For genes involved in epigenetic and transcriptional regulation, sequence-base analyses of RNA levels (RNAseq), chromatin accessibility (Assay for Transposase-Accessible Chromatin With Sequencing [ATACseq]), and protein-chromatin interactions (ChIPseq) assays can provide information about the functional consequences of variants. Moreover, deciphering the effects of LoF variants on these parameters provides a benchmark for assessing pathogenicity of missense variants. The identification of differences in the disrupted genes and pathways by variants in the same gene may also address mechanisms for differing clinical phenotypes. Applying these strategies, we assayed the epigenetic and transcriptional effect in induced pluripotent stem cells (iPSCs) carrying *SMAD2* LoF and missense variants identified in CHD probands.

The SMAD (suppressors of mothers against decapentaplegic) protein family encodes transcription factors with broad and crucial roles in early human embryonic development.[9–11] SMADs are grouped into 3 subtypes: receptor regulated (R-SMADs), common partner (co-SMADs), and inhibitory (I-SMADs).[9] The TGF-beta (transforming growth factor beta) superfamily of ligands comprises >30 proteins, including nodal, activin, and BMP bone morphogenetic proteins).[9–17] TGF-beta ligand activation of type I and type II serine/threonine kinase receptors induces phosphorylation of R-SMADs, SMAD2, and SMAD3 and promotes oligomerization of co-SMAD SMAD4. The complex then is translocated to the nucleus to promote transcriptional activation of genes essential for human developmental processes.[9–19]

SMAD proteins have 2 highly conserved domains, an N-terminal MH1 (mad homology 1) domain and a C-terminal MH2 domain, connected by a central proline-rich linker region.[20–23] The MH1 domain directs DNA binding by all R-SMADs, except SMAD2,

which contains inserted sequences within the MH1 β hairpin that prevent DNA binding.[21–23] The MH2 domain mediates interactions among SMAD 2, 3, and 4 and transcription factors[21,23] and forms complexes that bind both accessible (open chromatin) and nonaccessible (closed chromatin) DNA.[24–28] When associated with chromatin remodeler proteins, SMAD complexes bind open chromatin and promote transcription.[25–29]

SMAD2/3 is known to interact with transcription factors TEAD (transcriptional enhanced associate domain 3/4) and OCT4 (octamer-binding transcription factor 4) to form the TSO (TEAD, SMAD2/3, OCT4) complex.[30] TSO complex bound in proximity to FOXH1 (forkhead box protein H1) binding sites acts as an enhancer switch to modulate gene expression for early development and mesodermal cell fates.[30,31] Additionally, SMAD2/3 interactions with regulatory binding motifs for pluripotent factors OCT4, SOX2 (SRY-box 2), and NANOG (homeobox protein NANOG) play a pivotal role in specifying cellular fate and lineages.[30,32–34] SMAD2/3 also engages with a diverse array of transcription factors, including ETS, SP1 (specificity protein 1), and the AP1 (activator protein 1) family members, that participate in TGF-beta signaling and regulate stem cell renewal, apoptosis, and growth.[12,35–41] Both SP1 and AP1 transcription factors are involved in processes that regulate cardiac function as well as human embryonic stem cell cardiomyocyte (CM) differentiation and development.[12,42,43]

We identified de novo heterozygous *SMAD2* LoF and rare missense variants in pediatric patients with complex CHD.[4–6,44] *SMAD2* missense variants have also been found in adult patients with arterial aneurysms and dissections, with or without connective tissue abnormalities but not CHD.[44–49] A *SMAD2* LoF variant has been found in 1 patient with CHD and abnormal pulmonary venous return.[44]

We report analyses of the transcriptional and epigenetic effects of *SMAD2* variants identified by exome sequencing of 11 000 CHD probands participating in the Pediatric Cardiac Genomics Consortium (PCGC). Among 30 *SMAD2* variants, we engineered heterozygous and homozygous LoF and 2 missense *SMAD2* variants into isogenic iPSCs. By comparative analyses of RNAseq and ATACseq with published SMAD2/3 ChIPseq data, we demonstrate that *SMAD2* haploinsufficiency alters at least five transcription factor interactions with their cognate DNA binding sites. Recognizing that SMADs regulate pluripotency and lineage decisions, we determined the potential for mutant iPSCs to differentiate into beating CMs. Our analyses provide functional evidence that these LoF and missense variants likely cause CHD and contribute insights into transcriptional pathways disrupted by insufficient levels and inappropriate activities of SMAD2.

METHODS

All data and materials are publicly available, as detailed in Data S1.

PCGC Study Cohort

Participants with CHD (n=11 336) were recruited for the CHD GENES trial (Congenital Heart Disease Network Study of the PCGC; ClinicalTrials.gov identifier NCT01196182) and the DNA Biorepository of the Single Ventricle Reconstruction trial after approval from institutional review boards as previously described.[50,51] All participants or their parents provided written informed consent. Cardiac and extracardiac phenotypes, obtained from medical records and family interview are maintained in the PCGC datahub.

Exome Sequencing and Variant Filtering

DNA was extracted from blood or saliva samples and sequenced as previously described.[5,52–54] Sequence reads were mapped to the reference genome (hg38), and further processed using the Genome Analysis Toolkit Best Practices workflows as previously described.[5,52–54] Single nucleotide variants and small indels were called with Genome Analysis Toolkit HaplotypeCaller. The resulting variant call file was annotated using SNPEFF and ANNOVAR.[55,56] De novo variants were independently called and filtered for quality using the Trio Denovo program and a custom pipeline, which have been shown to yield a specificity of 96.3%.[2] Candidate LoF heterozygous variants were filtered for rarity (1000 genomes cohort allele frequency ≤0.001) and quality as previously described.[5] Genome aggregation database v3 (gnomAD[57]) cohort data were accessed on July 23, 2023.

CRISPR Gene Editing and Mutation Confirmation

SMAD2 variants were introduced into PGP1 (personal genome project 1) iPSCs (passage range 65–77). *SMAD2* LoF iPSCs were generated by nonhomologous end joining by coelectroporation using 2 μg of both plasmids encoding Cas9 (PX459v2; Addgene) and plasmid guide RNA (2 μg total), using a stem cell 4-D core nucleofector unit (Lonza). Heterozygous missense variants, $SMAD2^{+/W274C}$ and $SMAD2^{+/R114C}$, were introduced in human iPSCs using homology directed repair by nucleofection of 2 μg for Cas9 (PX459v2 from Addgene), plasmid guide RNA, and a single-stranded oligonucleotide repair template. Plasmid-containing clones were selected using puromycin (Gibco), expanded and genotyped as previously described.

Gene-edited cells were subcloned twice and validated by Sanger sequencing and next generation sequencing. Briefly, subcloning involved dissociation

of iPSCs by pipetting and filtering through a 60-μm strainer, plating on a 15-cm dish containing MTESR (Stemcell Technologies)+Rock inhibitor (10 nmol/L) (R&D Systems), and colony growth to 300 cells. Individual clones were picked and placed into separate wells of a 96-well plate, grown to 85% confluency, and then processed by polymerase chain reaction-amplification and sequencing. To ensure clone purity, initial iPSCs clones were subjected to an additional subcloning, after which polymerase chain reaction amplified fragments were studied by Sanger and next generation sequencing. Sequences was analyzed using DNA-star software, and Integrated Genomics Viewer. Two independent clones were created for each genotype (*SMAD2*$^{-/-}$: *SMAD2*$^{P135/P135*}$, *SMAD2*$^{P135/K157*}$; *SMAD2*$^{+/-}$: *SMAD2*$^{+/P135*}$; *SMAD2*$^{+/W274C}$; and *SMAD2*$^{+/R114C}$). For each experiment, we report data from 2 independent lines for each *SMAD2* genotype and at least 2 technical replicates for each sample.

Western Blots

iPSCs were sonicated for 2 minutes in 6×16 mm AFA microtubes (Covaris E210 focused ultrasonicator; duty factor 5%, and 200 Cycles/burst at 4 °C) and lysed using RIPA lysis buffer (Thermoscientific). Protein concentration was determined using a BCA Protein Assay Kit (Pierce) and NanoDrop 2000/2000c Spectrophotometer. For immunoblots, samples containing approximately 30 μg of denatured protein were reacted with, in NuPAGE Reducing Buffer (Invitrogen) and LDS Sample Buffer (Pierce). After denaturation, samples were loaded onto a Novex 4% to 20% Tris-Glycine Mini Gel cassettes (Thermo Scientific). Gels were transferred overnight at 16 °C at 100 mA onto a PVDF membrane and blocked for 1 hour. We used primary human/mouse SMAD2/3 (R&D Systems, cat no. AF3797) and rabbit monoclonal SMAD2 antibodies (Cell Signaling, cat no. 5339S) at a 1:1000 dilution for 1 hour. Beta actin (Thermo Scientific, cat no. MA5-15739) at a 1:5000 dilution and GAPDH monoclonal antibody (Thermo Scientific, cat no. MA5-15738-HRP) at a 1:2000 dilution was used to assess loading. We used secondary goat antirabbit IgG HRP antibody (Thermo Scientific, cat no. A21207) and goat antimouse IgG peroxidase-labeled antibody (cat. no. 115–035-003; Jackson ImmunoResearch Laboratories, Inc.). Western blots bands were visualized using a ChemiDocTM XRS+ and ImageLab v3.0.

Bulk RNAseq Library Prep and Analysis

iPSC RNA was purified using Trizol (Life Technologies) with RNA integrity numbers >8. Nextera libraries (Illumina) were prepared, and samples pooled for sequencing using the Illumina NextSeq500 platform with 4 lanes (1 flow cell). Data were combined into a single fastq file, reads aligned to the hg38 reference genome using STAR, and mitochondrial and duplicate reads discarded using Samtools and Picard's MarkDuplicates, respectively. Aligned reads were quantitated by feature counts and normalized. Gene expression for SMAD components (SMAD2, SMAD3, and SMAD4) were calculated as reads per gene per million aligned reads (rpkm). Significance (defined as $P<0.05$) was assessed using Student's *t* test for comparison between data sets. All *P* values were adjusted by Bonferroni correction. Differential expressed genes were identified using DESeq2. Comparison between data sets were analyzed using the Wald test, with significance defined as $P<0.05$. All adjusted *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure.

iPSC-Derived CM Differentiation

iPSCs, maintained in feeder-free conditions with MTESR (Stemcell Technologies) media, were differentiated to the CMs by sequential targeting of the WNT pathway.[58,59] CMs were purified using glucose metabolic selection and studied on days 30 to 40 after initiation of differentiation except where indicated otherwise.[58,59]

Immunofluorescence

iPSC-CMs were seeded onto sterile, acid-treated, 18-mm #1.5 glass coverslips in 24-well plates, fixed in 2% paraformaldehyde in PBS for 10 minutes, and washed with PBS. Permeabilization was accomplished by placing cells for 5 minutes in PBS containing 0.1% Triton X-100 at 37 °C. Cells were incubated for 1 hour at room temperature with primary rabbit polyclonal cardiac troponin T (Abcam, cat. no. ab45932) and mouse monoclonal α–actinin (Sigma-Aldrich, cat. no. a7811) antibodies, and then followed by 3 washes to remove unbound antibody. Cells were then incubated for 1 hour with secondary goat antirabbit fluorescein isothiocyanate (Jackson Laboratories, cat. no. 656111) and goat antimouse rhodamine (Jackson Laboratories, cat no. 31663) antibodies. DNA was stained with 4,6-diamidino-2-phenylindole (Sigma, cat no. D9564) at a 1:5000 dilution for a 2-minute incubation. Antibody buffer without primary antibody was used to assess nonspecific binding of secondary antibodies (fluorescein isothiocyanate and rhodamine). Cells were imaged in a Yokogawa CSU-W1 spinning disk scan head with a 50-μm pinhole disk mounted on a Nikon Ti inverted microscope (Nikon Ti), equipped with a Nikon motorized stage with a Physik Instrument piezo Z motor, a Plan Apo Lambda 100×/1.45 DIC objective, and a Andor Zyla 4.2 plus sCMOS camera. Images were acquired using NIS Elements AR 5.02. Signal from fluorescein isothiocyanate and rhodamine channels was

collected using a Chroma ET 525/36 and ET 605/52 emission filters, respectively. Representative immunofluorescent images were selected to that provide the best sarcomere phenotype of each cell line.

For live cell imaging analysis, iPSC-CMs were differentiated in 6-well plates (Corning) using RPMI-1640 media containing B27 supplement. Live cell images were collected using a Keyence BZ-X710 microscope at room temperature using a 60× objective.

Functional Assays of Day 30 iPSC-CMs

Sarcomere contraction of wild-type (WT) and SMAD2 mutant iPSC-CMs, transfected with GFP (green fluorescent protein)-actinin lentivirus to enable high fidelity tracking of sarcomere function, was assessed using SarcTrack.[60] Five-second videos were acquired on 3 separate wells of differentiation day 30 iPSC-CMs, paced at 1 Hz, 12 volts per 0.5 milliseconds and analyzed using SarcTrack. Significance (defined as $P<0.05$) was assessed using the Student *t* test. Multiple comparisons between genotypes were analyzed using 1-way ANOVA with post hoc Bonferroni correction, with significance defined as $P<0.05$.

Assay for Transposase-Accessible Chromatin With Sequencing

Nuclei were isolated from approximately 50 000 cells, treated with Tn5 transposase (Nextera DNA Sample Prep Kit, Illumina), and DNA isolated. The resultant fragmented DNA was amplified using bar-coded polymerase chain reaction primers,[61] and libraries were pooled and sequenced (Illumina Next-Seq) to a depth of 100 million reads per sample. Reads were aligned to the hg38 reference genome using BWA-MEM and peaks were called using Hypergeometric Optimization of Motif Enrichment (v4.10.3; http://homer.ucsd.edu/homer/index.html). Analyses of ATACseq peaks, differential peaks, and transcription binding site motif enrichment were studied using Hypergeometric Optimization of Motif Enrichment (v4.10.3).

Gene Ontology Analysis

Gene ontology annotation was derived using the R package clusterProfiler.[62] Genes were classified by Gene Ontology annotation based on biological process, molecular function or cellular component. Genes expressed in WT PGP1 iPSCs were used as a background and default setting were used for other parameters. Bonferroni correction for multiple testing was used to determine significance thresholds.

Statistical Analysis

Single comparisons were analyzed by using the Student *t* test, with significance defined as $P<0.05$. For functional assays of day 30 iPSC-CMs, multiple comparisons between genotypes were analyzed using 1-way ANOVA with post hoc Bonferroni correction, with significance defined as $P<0.05$.

RESULTS

SMAD2 Variants in CHD Probands

Exome sequencing of 11 336 PCGC CHD probands[4–6] identified 5 rare LoF variants in *SMAD2* (Figure 1, Table S1). Parental analyses indicated 1 de novo

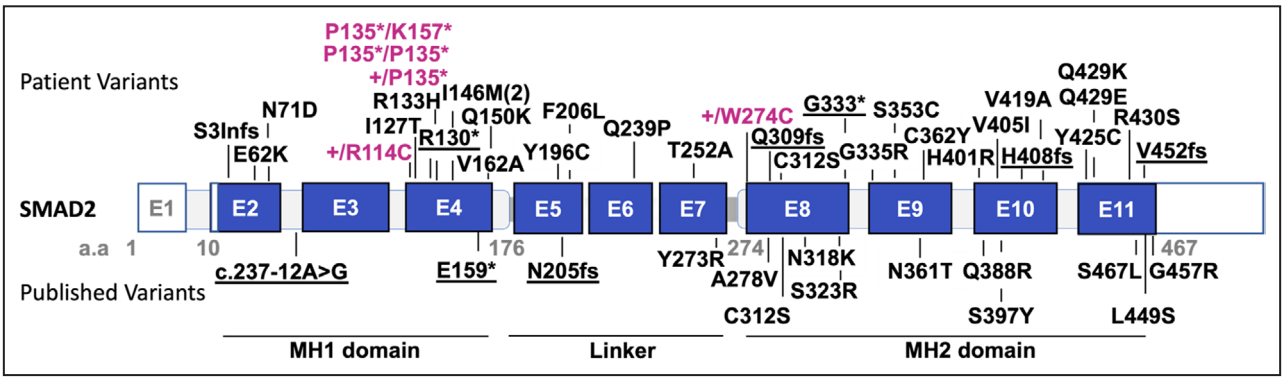

Figure 1. *SMAD2* variants from participants with congenital heart disease.
The *SMAD2* gene consists of the MH1 and MH2 domains, connected by a linker region. The MH1 domain primarily encodes DNA binding for SMAD transcription factors, with the exception of *SMAD2*, which contains a sequence insert that prevents direct binding to DNA. The MH2 domain mediates protein–protein interactions with co-SMADs and other transcription factors. (**Top**) Five loss-of-function (underlined) and 24 missense variants were identified in this cohort of participants with CHD. Five variants were modeled in iPSCs using CRISPR/Cas9 gene editing tools (magenta). *SMAD2* variants identified in participants with CHD were prevalent in exon 4 (binomial $P=0.05$). (**Bottom**) Three loss-of-function (underlined) and 11 missense variants have been reported in earlier studies; 10 of these missense variants are located at the MH2 domain of the *SMAD2* gene ($P=0.01$). CHD indicates congenital heart disease; and MH, MAD homology.

(p.G333*), 1 inherited (p.Val452fs), and 3 unclassified variants due to absence of parental data.[4–6] Three CHD probands had heterotaxy (cardiac or abdominal) syndrome with cardiac phenotypes of double-outlet right ventricle, atrioventricular canal, hypoplastic left heart syndrome, and transposition of the great arteries. Two probands without heterotaxy had comparable CHD phenotypes and aortic coarctation. CHD probands with *SMAD2* LoF variants also had vascular anomalies involving arterial (pulmonary) abnormalities, venous anomalies of the superior vena cava, or hepatic and systemic veins. Previous studies of 5 patients with *SMAD2* LOF variants reported comparable CHD phenotypes in 2 and isolated (aortic or thoracic) aneurysms in 3.[44,46]

We also identified rare *SMAD2* missense variants in 25 PCGC CHD probands (Figure 1, Table S1). Analyses of parental sequences, available for 12 probands, indicated that 3 were de novo and 9 were rare (allele frequency<1.0e-5 or absent from the gnomAD database[57]) and inherited. One missense variant, p.Ile146Met, occurred in 2 probands. These data indicate the frequency of rare *SMAD2* missense variants among CHD probands is ~0.002 (25/11 336 CHD probands), approximately twice the observed frequency of rare missense variants reported in a population-based gnomAD database (n=114/125 179; P=0.0009).

CHD phenotypes associated with *SMAD2* missense variants were similar to LoF variants (Figure 1, Table S1). Eight probands had heterotaxy (cardiac or abdominal) including doublet-outlet right ventricle, atrioventricular canal, tetralogy of Fallot, hypoplastic left or right ventricles, and vascular anomalies. Seventeen probands without heterotaxy syndrome had comparable CHD phenotypes, but only 11 had vascular anomalies involving the aorta or pulmonary arteries and pulmonary or systemic veins. *SMAD2* missense variants in CHD probands were clustered in exon 4 (encoding part of the MH1 domain) more than other exons (P=0.05, Figure 1). Notably, there were more heterotaxy cases in probands with a SMAD2 LoF or missense variant than probands without a SMAD2 variant (P=1.35e-07, Table S1).

Among 17 previously reported individuals with *SMAD2* missense variants, 15 had arterial abnormalities and aneurysms with and without connective tissue disease.[44–49] Notably, 1 individual had CHD without arterial abnormalities. Ten of the *SMAD2* missense variants in the 17 patients with arterial anomalies localize primarily to the *SMAD2* MH2 domain (compared with CHD-associated missense variants, P=0.01, Figure S1).

We used the deep model predictive learning tool AlphaMissense[63] to evaluate the potential pathogenicity of 35 unique *SMAD2* missense variants (Figure 1, Table S1). Among these, 26 were classified as pathogenic, 6 as benign, and 3 were unassigned. Parallel analyses using Combined Annotation Dependent Depletion[64] and meta-analytic support vector machine[65] predictive algorithms were concordant for 28 of 32 variants classified by AlphaMissense (Table S1).

iPSC Models Carrying *SMAD2* Variants

We introduced *SMAD2* variants into human iPSC line PGP1 using CRISPR/Cas9 gene technologies (METHODS) and generated 2 independent isogenic cloned cell lines for each sequence-confirmed genotypes *SMAD2*$^{-/-}$, *SMAD2*$^{+/-}$, *SMAD2*$^{+/W274C}$, and *SMAD2*$^{+/R114C}$ (Figure 1, Figure S2A through S2C, Table S2).

SMAD2 expression in WT and mutated iPSCs, characterized by RNAseq (Figure S2D, Table S3), indicated that *SMAD2*$^{+/-}$ (2±0.2 RPKM) and *SMAD2*$^{-/-}$ (1±0.2 RPKM) iPSCs had 40% and 60% mean RNA levels compared with WT (mean=3.4±0.1 RPKM). Western blot analyses confirmed these data: *SMAD2*$^{+/-}$ compared with WT iPSCs had 63%±6% lower protein levels, whereas protein levels were profoundly diminished in *SMAD2*$^{-/-}$ iPSCs (Figure S2E and S2F). RNA levels in *SMAD2*$^{+/W274C}$ and *SMAD2*$^{+/R114C}$ were comparable to WT, but protein levels were reduced by 20%±6% and 44%±14% in *SMAD2*$^{+/R114C}$ and *SMAD2*$^{+/W274C}$ iPSCs, respectively.

Because *SMAD2* associates with components *SMAD3* and *SMAD4*,[9,17,20] we examined levels of these transcripts in the mutant iPSCs to WT (Figure S2G and S2H). *SMAD3* RNA expression in *SMAD2*$^{+/-}$ (3±0.2 RPKM), *SMAD2*$^{-/-}$, (3±0.4 RPKM), *SMAD2*$^{+/W274C}$ (3±0.1 RPKM), and *SMAD2*$^{+/R114C}$ iPSCs (3±0.5 RPKM) had ~25% lower mean RNA levels compared with WT (4±0.3 RPKM). Whereas *SMAD4* RNA expression was increased in *SMAD2*$^{+/W274C}$ iPSCs (12±1 RPKM), no change was observed in *SMAD2*$^{+/-}$, *SMAD2*$^{-/-}$, *SMAD2*$^{+/R114C}$ iPSCs to WT (10±0.6 RPKM). The consequences of *SMAD2* variants appeared to relate to attenuated SMAD levels and functions.

SMAD2 Variants Influence Cardiomyocyte Differentiation

As prior studies implicate SMADs function in mesodermal cell fate determination,[30,66–69] we assessed the potential for CM differentiation in *SMAD2* mutant iPSCs. Using our standard protocol,[58,59] *SMAD2*$^{+/-}$ and *SMAD2*$^{+/W274C}$ iPSCs differentiated into beating CMs with immunofluorescent cardiac troponin T staining of sarcomeres (Figure 2A, Table S2). Contractility of WT and *SMAD2* mutant iPSC-CMs were additionally monitored by live image analysis (Video S1 through S3).

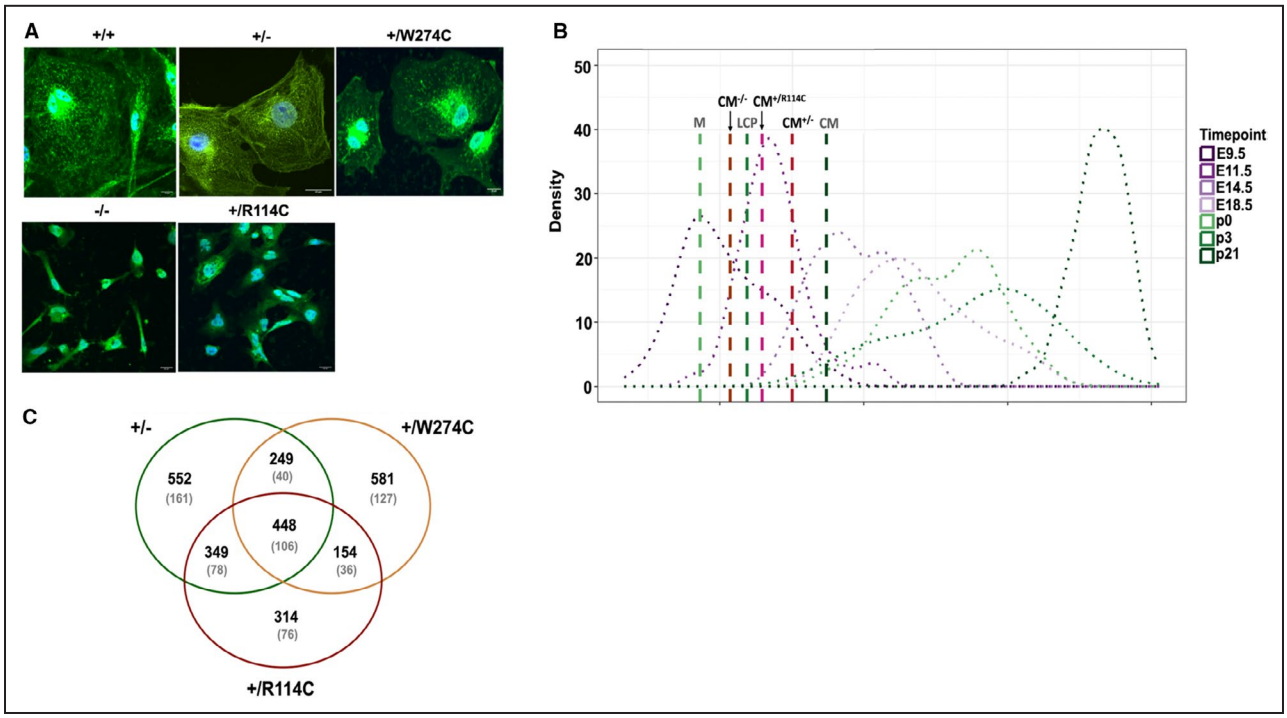

Figure 2. Cardiac maturation and transcriptional profiles in *SMAD2* haploinsufficient and missense iPSCs.
A, Representative images of WT, *SMAD2*$^{+/-}$, *SMAD2*$^{+/W274C}$, *SMAD2*$^{-/-}$, and *SMAD2*$^{+/R114C}$ iPSC-CMs stained with cardiac troponin T antibody (green) and 4,6-diamidino-2-phenylindole for nuclei (blue). A representative for each cell type is presented. Magnification, 40×, scale bar: 20 μmol/L. Sarcomeres were observed in WT, *SMAD2*$^{+/-}$, and *SMAD2*$^{+/W274C}$ iPSC-CMs, but absent in *SMAD2*$^{-/-}$ and *SMAD2*$^{+/R114C}$ iPSC-CMs. **B**, E9.5-p21 mouse ventricular cardiomyocyte transcriptional profiles displayed as dashed lines in density plot as previously described.[70] Vertical lines represent the transcriptional profiles of wildtype and *SMAD2*$^{-/-}$ iPSCs differentiation in the mesoderm stage (day=4), LCP (day=8) and CM (day=30) as determined by bulk RNAseq. Wild-type iPS-CMs (CM) correspond to mouse E14.5–E18.5 cardiomyocytes. *SMAD2*$^{+/-}$ iPSC-CM (CM$^{+/-}$) maturation is slightly delayed compared with wild-type iPSC-CMs. *SMAD2*$^{-/-}$ and *SMAD2*$^{+/R114C}$ iPS-CMs correspond to mouse E9.5–E11.5 cardiomyocytes and do not progress beyond mesoderm and late cardiac progenitor stages of development. **C**, Overlap of DEGs with minimum $\log_2$fold change|1| and $P<0.05$ in *SMAD2*$^{+/-}$, *SMAD2*$^{+W274C}$, and *SMAD2*$^{+/R114C}$ iPSCs as compared with wild-type iPSCs. All adjusted *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure. The number of DEGs near a SMAD2/3 chromatin immunoprecipitation peak[24] from human embryonic stem cells are indicated in gray. Data were collected from 2 independent cell lines for each genotype with technical replicates for selected lines. Total cell lines analyzed for each genotype in (**B**) and (**C**) *SMAD2*$^{+/+}$ (n=3), *SMAD2*$^{-/-}$ (n=4), *SMAD2*$^{+/-}$ (n=4), *SMAD2*$^{+/W274C}$ (n=3), and *SMAD2*$^{+/R114C}$ (n=3). CM indicates cardiomyocyte; DEG, differentially expressed gene; iPSC, induced pluripotent stem cell; LCP, late cardiac progenitor; M, mesoderm; and WT, wild-type.

We quantified sarcomere shortening in *SMAD2*$^{+/-}$ compared with WT iPSC-CMs by live image analysis, using GFP-labeled sarcomeres measurements across the contractile cycle. In comparison to WT, *SMAD2*$^{+/-}$ iPSC-CMs had a higher percentage fraction of sarcomere shortening (Videos S4 and S5 and Figure S3A). From bulk-RNA sequencing data, we found slightly reduced levels of *ACTC1 and MYH7* in *SMAD2*$^{+/-}$ iPSC-CMs compared with WT (Figure S3B, Table S4), supporting the observations that cardiac function and contractility were compromised in *SMAD2*$^{+/-}$ iPSC-CMs.

By contrast, differentiation of the *SMAD2*$^{+/R114C}$ and *SMAD2*$^{-/-}$ iPSCs failed to generate sarcomeres and did not contract (Figure 2A, Table S2). Suspecting this reflected stalled differentiation, we assessed a developmental index of progressive mouse embryonic cardiac transcriptional gene levels[70] from e9.5-p21[70] (Figure 2B). Comparison of serial RNA expression profiles of WT mesoderm (day=4), late cardiac progenitor (day=8), and CM (day=30) indicated that the development of *SMAD2*$^{-/-}$ and SMAD2$^{+/R114C}$ iPSCs failed to progress beyond the mesoderm stage (Figure 2B, Figure S3C). In contrast, SMAD2$^{+/-}$ iPSC-CMs expression profiles showed maturation beyond the late cardiac progenitor stage, albeit slightly delayed compared with WT iPSC-CMs.

Transcriptome Profiles in *SMAD2* Variant iPSCs

We profiled 2 independent biological replicates for each genotype of iPSCs. Principle component analysis of RNAseq analyses confirmed transcriptional similarities (Figure S4A). Comparative analyses of mean RNA expression between WT and mutant lines revealed 1599

Table 1. Genes With Altered Expression in *SMAD2* Mutant iPSCs Compared With WT

Genotype	DEGs*	DEGs near a SMAD2/3 peak†	Upregulated in mutant iPSCs	Downregulated in mutant iPSCs
+/–	1599	385	923	676
+/W274C	1432	309	776	656
+/R114C	1265	296	630	635

DEG indicates differentially expressed gene; iPSC, induced pluripotent stem cell; and WT, wild-type.

*Gene expression changes in *SMAD2* mutant iPSCs were compared with wild-type iPSCs and those with absolute $\log_2$fold change of at least 1, and $P<0.05$ were determined to be differentially expressed.

†Differentially expressed genes near a SMAD2/3 chromatin immunoprecipitation peak. Number of peaks analyzed in SMAD2/3 data set=10 566 peaks, number of genes associated with a SMAD2/3 peak=7087 genes, number of genes associated with a SMAD2/3 peak expressed in personal genome project 1 iPS cells (>1 read per kilobase per million mapped reads)=3899 genes.

differentially expressed genes in *SMAD2*$^{+/-}$, among which 1046 were also dysregulated in *SMAD2*$^{+/W274C}$ or *SMAD2*$^{+/R114C}$ iPSCs (Figure 2C, Table 1, Figure S4B, Table S3). These data implied these missense variants had similar functional effects to LoF variants.

SMAD2-mutant iPSCs had reduced levels of target genes in the TGF-beta/BMP pathway[13,14,17,71,72] that regulate SMAD2/3 signaling (Table S3 and S5), including *BMP2*, *EGR1*, *CER1*, *GDF3*, *HTRA1*, and *SFRP2*.[71] RNA levels of *NODAL*, *LEFTY1*, and *LEFTY2*, components of the activin/nodal signaling pathway,[73] were also significantly reduced in *SMAD2*$^{+/-}$ and *SMAD2*$^{+/R114C}$ iPSCs (Figure S4C, Table S3). Furthermore, in *SMAD2*$^{+/-}$ iPSCs, RNA levels of the pluripotent factor *TEAD4* increased 2-fold compared with WT. However, RNA levels of *OCT4, SOX2, TEAD3, and NANOG* were comparable in mutant iPSCs to WT, indicating that a reduction in TGF-Beta/SMAD signaling did not affect the pluripotent state of these cells.

Next, we assessed published SMAD2/3 ChIP-seq data from embryonic stem cells (Gene Expression Omnibus accession number: GSE29422),[24] to identify direct transcriptional SMAD targets (Figure 2C, Table 1). SMAD2/3 ChIP-seq peaks occupied 24% (n=385/1599) differentially expressed genes in *SMAD2*$^{+/-}$ iPSCs, implying these direct targets were affected by *SMAD2* haploinsufficiency. Ten of these 385 genes were previously identified as CHD genes, including *ARID1A*, *BCOR*, *COL5A1*, *COL5A2*, *FGF8*, *FGFR1*, *GLI3*, *NODAL*, *LEFTY2, and TBX1*.[74] One-hundred

Table 2. - SMAD2/3-Bound Open and Closed Chromatin in *SMAD2*-Mutant iPSCs Compared With WT

	Genotypes			
	+/+	+/–	+/W274C	+/R114C
Open chromatin*				
ATAC peaks	147 572	144 934	171 442	201 525
Overlapping SMAD2/3 binding sites	5239	4672	4864	5498
Associated genes†	3975	3704	3786	4054
Associated DEGs‡		194	146	227
Total differential *SMAD*-ATAC peaks‡		3299	3104	1022
Reduced peaks		3191	2932	918
Enhanced peaks		108	171	401
Associated genes§		1815	1738	708
Associated DEGs§		183	128	52
Closed chromatin*				
SMAD2 binding sites without ATAC peaks	5327	5894	5702	5068
Associated genes (>1 RPKM in WT iPSCs)	3975	3704	4105	3818
Associated DEGs‡		221	181	193

ATACseq was performed in WT and *SMAD2* mutant iPSCs. These regions were overlapped with 10 566 published SMAD2/3 peak. Genes were near a SMAD2/3 peak, of which genes were expressed >1 RPKM in personal genome project 1 iPSCs. SMAD2/3 chromatin immunoprecipitation peaks within regions of ATACseq peaks of iPSCs were referred to as open chromatin peaks, and SMAD2/3 ChIPseq peaks occurred in both open chromatin (identified by an ATACseq peak) and closed chromatin (ie, no ATACseq peak).

ATACseq indicates Assay for Transposase-Accessible Chromatin With Sequencing; DEG, differentially expressed gene; HOMER, Hypergeometric Optimization of Motif Enrichment; iPSC, induced pluripotent stem cell; RPKM, reads per kilobase per million mapped reads; and WT, wild-type.

*SMAD2/3 binding sites without ATAC peaks.

†HOMER analysis was used to assess genes (nondifferential and differentially expressed) near a *SMAD*-ATAC peak.

‡SMAD2/3 bound open chromatin peaks in mutant lines were compared with WT. Differential peaks were calculated at a 1.5-fold change and *P* value <1e-4.

§HOMER analysis was used to assess genes (nondifferential or differentially expressed) near a reduced or enhanced -SMAD2/3 bound open chromatin peak.

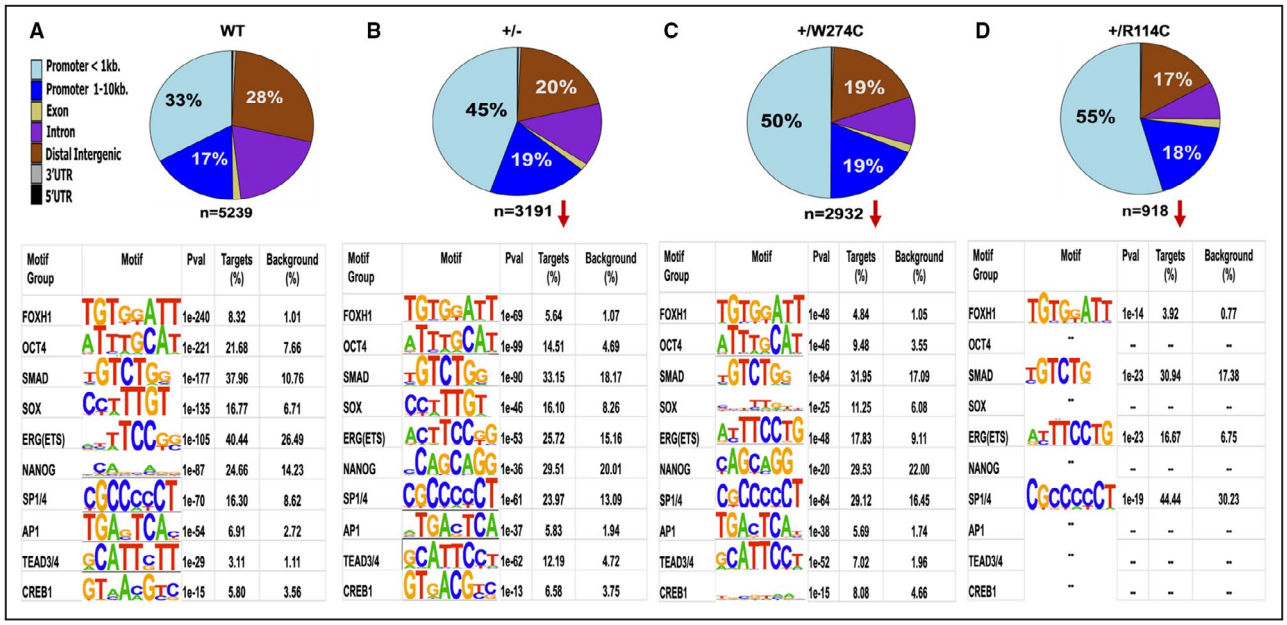

Figure 3. Changes in SMAD-bound open chromatin peaks due to *SMAD2* variants.
The location of SMAD-bound open chromatin peaks was characterized with respect to gene bodies in wild-type and *SMAD2* variant cells and noted as percentage of total peaks (**top**). DNA-binding motif enrichment was performed using Hypergeometric Optimization of Motif Enrichment analysis (**bottom**). **A**, Wild-type iPSCs have half of chromatin peaks in promoter regions, and are enriched for SMAD2/3/4, FOXH1, OCT, SOX, ETS, NANOG, SP1/4, TEAD3/4, AP1 and CREB1 motifs. **B**, *SMAD2*$^{+/-}$ iPSCs have a reduction in 3191 of the 5239 SMAD-bound open chromatin peaks in wildtype cells, with differential enrichment of ETS, NANOG, TEAD3/4 and CREB1 motif sequences compared with motifs observed in wild-type iPSCs. **C**, *SMAD2*$^{+/W274C}$ iPSCs have a reduction in 2932 SMAD-bound open chromatin peaks in wild-type cells with altered motif sequences of SOX, ETS, CREB1 compared with motifs observed in wild-type and *SMAD2*$^{+/-}$ iPSCs. The motifs for NANOG and TEAD3/4 were similar in *SMAD2*$^{+/-}$ and *SMAD2*$^{+/W274C}$ iPSCs, but differed from wild-type iPSCs. **D**, *SMAD2*$^{+/R114C}$ iPSCs showed fewer reductions in SMAD-bound open chromatin peaks (n=918) than *SMAD2*$^{+/-}$ and *SMAD2*$^{+/W274C}$ iPSCs. In comparison to wild-type and *SMAD2* mutant cells, reduced peaks showed no enrichment of OCT4, SOX, NANOG, AP1, TEAD3/4, and CREB1 motifs, but changes were detected in SMAD and ETS motifs. Data were collected from 2 independent cell lines for each genotype and as technical replicates. Total cell lines analyzed: *SMAD2*$^{+/+}$ (n=4), *SMAD2*$^{-/-}$ (n=4), *SMAD2*$^{+/-}$ (n=4), *SMAD2*$^{+/W274C}$ (n=4), and *SMAD2*$^{+/R114C}$(n=4). AP1 indicates activator protein 1; CREB1, cAMP response element binding protein 1; FOXH1, forkhead box protein H1; iPSC, induced pluripotent stem cell; NANOG, homeobox protein NANOG; OCT, octamer-binding transcription factor; *P* value <1e-4,; SOX, SRY-box 2; SP1/4, specificity protein 1/4; and TEAD3/4, transcriptional enhanced associate domain 3/4.

and six of these 385 direct targets genes are also differentially expressed in both *SMAD2*$^{+/R114C}$ and *SMAD2*$^{+/W274C}$ iPSCs (Figure 2C). Gene Ontology enrichment analysis of these 106 genes (Figure 2C, Table S6) identified pattern specification processes (adjusted for Bonferroni testing, *P*=0.0005), inclusive of *BCOR*, *FGF8*, and *TBX1* genes. Together these data support the conclusion that heterozygous *SMAD2* missense variants W274C and R114C like LoF variants, disrupt common cardiac developmental processes and thereby cause CHD.

Transcription Factor Binding Sites Affected by *SMAD2* Variants

We probed whether *SMAD2* variants affected chromatin open/closed states using ATACseq combined with SMAD2/3 ChIPseq data from human embryonic stem cells[24] and identified changes in binding motifs using Hypergeometric Optimization of Motif Enrichment.[75] ATAC peak totals were similar in WT and *SMAD2*$^{+/-}$ iPSCs (~147 000) and lower than in missense iPSCs (~170 000 and 200 000) (Table 2). SMAD2/3 ChIPseq data[24] overlapped with ~5000 ATAC peaks in both open and closed chromatin observed in WT or mutant iPSC lines.

In WT iPSCs, ~50% of open chromatin regions bound by SMAD2/3 resided in promoter regions that were also enriched for binding motifs of interacting partners, including pluripotent embryonic factors OCT4, SOX, NANOG, and TEAD3/4, stem cell renewal marker ETS, and mesoderm regulator FOXH1 (Figure 3A, Figure S5A). Additional enriched binding motifs within these open chromatin ATAC peaks included SP1/4, AP1, and CREB1 (cAMP response element binding protein 1) transcription factors that interact with the SMAD2/3 complex and function as effectors of TGF-beta signaling.[12,35–39,76]

SMAD2/3-bound similar locations of open (Figure 3A) and closed chromatin in WT iPSCs (Table 2,

Figure S5). These sequences are predicted to bind the same transcription factors targeted by SMAD, albeit with differences in the binding motifs. Additionally, there was a greater proportion FOXH1 motifs (19.3% closed versus 8.3% open) and no enrichment for NANOG, TEAD3/4, SP1/4, or CREB1 motifs in closed chromatin, consistent with the undifferentiated state of WT iPSCs.

Over 95% of differential SMAD-ATAC peaks for each *SMAD2* mutant were reduced compared with WT iPSCs (Table 2, Figure 3A through 3D, Table S7) and predominantly resided in promoter regions (≥64%). Among all (~3000) reduced SMAD-ATAC peaks in *SMAD2*$^{+/-}$ and *SMAD2*$^{+/W274C}$ iPSCs compared with WT iPSCs, Hypergeometric Optimization of Motif Enrichment identified distinctive motifs for transcription factors. For example, motifs for ETS, NANOG, TEAD3/4, and CREB1 differed in *SMAD2*$^{+/-}$ and WT iPSCs (Figure 3A and 3B). The motifs for NANOG and TEAD3/4 were shared in *SMAD2*$^{+/-}$ and *SMAD2*$^{+/W274C}$ iPSCs while the CREB1 motifs differed between both mutant and WT iPSCs (Figure 3A through 3C). In addition, the SOX motif in *SMAD2*$^{+/W274C}$ differed from both *SMAD2*$^{+/-}$ and WT iPSCs. We also observed differential enhanced SMAD-ATAC peaks in *SMAD2*$^{+/-}$ (~100) and *SMAD2*$^{+/W274C}$ (~170) compared with WT iPSCs with 2 predominant motifs (Table 2, Figure S6A and S6B, Table S7); the FOXH1 motif differed in *SMAD2*$^{+/-}$ and the SOX motif differed from WT iPSCs in both mutant lines.

SMAD2$^{+/R114C}$ had the fewest reduced differential SMAD-ATAC peaks (~1000) in open chromatin compared with WT iPSCs and lacked enrichment for OCT4, SOX, NANOG, AP1, CREB1, and TEAD3/4 motifs found in WT. *SMAD2*$^{+/R114C}$ also lacked the distinctive motifs for these 6 transcription factors that were identified in other mutant iPSCs (Table 2, Figure 3A and 3D, Table S7). However, the ETS motif in *SMAD2*$^{+/R114C}$ was similar to other mutant lines and differed from WT iPSCs. *SMAD2*$^{+/R114C}$ had the most (~400) differential enhanced SMAD-ATAC peaks compared with WT, with cognate binding motifs for NANOG and TEAD3/4 (Table 2, Figure S6A through S6C, Table S7). The enhancement of SMAD-ATAC peaks was enriched for distinct motifs for SOX3 and FOXH1; the SOX3 motif differed from other mutant iPSCs, whereas the FOXH1 motif was unique to *SMAD2*$^{+/R114C}$ iPSCs.

Transcriptional Responses to Changes in Chromatin Accessibility and Association With CHD in *SMAD2* Variant iPSCs

We explored whether the reduction of SMAD-bound open chromatin ATAC peaks correlated with differentially expressed genes in mutant compared with WT iPSCs, and whether these genes were relevant to CHD (Figures 2C, 4A through 4F, Tables 1 and 2). In *SMAD2*$^{+/-}$ iPSCs, 183 of 385 direct and dysregulated

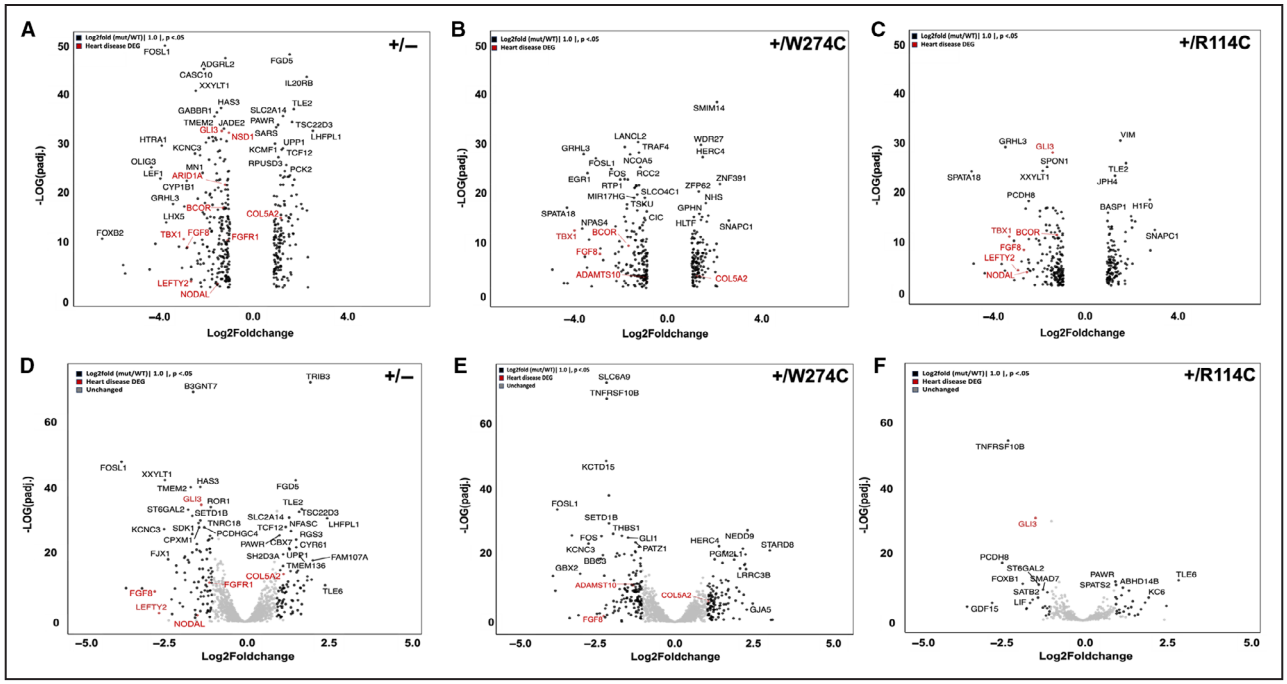

Figure 4. Differential expressed direct targets of *SMAD2*.
A through **C**, Differentially expressed genes near a SMAD2/3 chromatin immunoprecipitation peak in (**A**) *SMAD2*$^{+/-}$, (**B**) *SMAD2*$^{+/W274C}$, and (**C**) *SMAD2*$^{+/R114C}$ iPSCs. **D** through **F**, Differential expressed genes near a reduced SMAD-bound open chromatin peak in (**D**) *SMAD2*$^{+/-}$, (**E**) *SMAD2*$^{+/W274C}$, and (**F**) *SMAD2*$^{+/R114C}$ iPSCs. Differential expression was considered for log2 fold change |1| and P>0.05. Genes known to cause CHD are highlighted in red. DEG indicates differentially expressed gene; and iPSC, induced pluripotent stem cell.

SMAD targets resided in reduced SMAD-bound open chromatin ATAC peaks and included 6 CHD genes (*FGF8*, *LEFTY2*, *NODAL*, *GLI3*, *FGFR1*, and *COL5A2*).[74] With the exception of upregulation of *COL5A2*, the expression of these genes were significantly reduced in *SMAD2*$^{+/-}$ iPSCs (Figure 4D).

Similarly, direct and dysregulated SMAD targets in *SMAD2*$^{+/W274C}$ and *SMAD2*$^{+/R114C}$ iPSCs resided in reduced SMAD-bound open chromatin ATAC peaks and were associated with CHD genes (Figures 2C and 4B, 4C, 4E, 4F, Tables 1 and 2). The genes associated with *SMAD2*$^{+/W274C}$ epigenetic changes included *FGF8*, *ADAMTS10*, and *COL5A2*.[74] Epigenetic changes associated with *SMAD2*$^{+/R114C}$ iPSCs involved fewer (n=52) direct and dysregulated SMAD target genes and were associated with only 1 CHD gene, *GLI3*[74] (Figure 4F).

DISCUSSION

We demonstrate that *SMAD2* LoF and missense variants, a previously identified cause of CHD, altered epigenetic and transcriptional processes that orchestrate the expression of genes required for heart development.[4,6,44–48] Analysis of exome sequences from ~11 000 patients with CHD identified 31 subjects with *SMAD2* rare inherited or de novo variants with a range of cardiac malformations and vascular anomalies. These variants were distributed across all 3 domains of the *SMAD2* gene, although more variants clustered in exon 4 (MH1 domain) than in other exons. Other *SMAD2* missense variants identified in adults with arterial aneurysms primarily localize to the MH2 domain.[44–48] Deep model predictive artificial intelligence technology, such as AlphaMissense,[63] a robust bioinformatic tool predicted 26 of 30 *SMAD2* variants as pathogenic, with high concordance to Combined Annotation Dependent Depletion[64] and meta-analytic support vector machine[65] algorithms.

We explored mechanisms that may cause distinct cardiovascular phenotypes using human iPSCs and sequence-base analysis to evaluate changes in RNA expression (RNA-seq) and chromatin accessibility containing SMAD2/3 binding sites (ATAC-seq and ChIP-seq). *SMAD2*$^{+/-}$ iPSCs had approximately 50% of WT protein levels, and like *SMAD2*$^{+/W274C}$ iPSCs, differentiated into CMs, albeit with compromised function. Further analysis showed *SMAD2*$^{+/-}$ iPSCs had above average sarcomere shortening. *SMAD2*-null cell lines produced little SMAD2 protein and, like *SMAD2*$^{+/R114C}$ iPSCs, failed to differentiate into CMs, highlighting the essential role of *SMAD2* in early cardiac development.

By integrating published ChIPseq data[24] with ATACseq analyses, we found that SMAD2/3 predominantly localizes to the promoter regions of both open and closed chromatin. Moreover, the binding motifs for key transcription factors (OCT4, SOX, ETS, AP1), but not FOXH1, differed in open and closed chromatin. Within open chromatin, *SMAD2* haploinsufficiency bound motifs that differed from WT iPSC for transcription factors involved in stem cell pluripotency (NANOG and TEAD) and TGF-beta signaling (ETS, AP1, and CREB1). These epigenetic changes dysregulated 183 direct SMAD target genes and subsequently resulted in an additional 202 dysregulated genes (Figure 4A and 4D). In addition to TGF-beta signaling genes, the dysregulated genes participate in pattern specification processes, and 10 are previously identified as CHD genes (*ARID1A*, *BCOR*, *COL5A1, COL5A2*, *FGF8*, *FGFR1*, *GLI3*, *NODAL, LEFTY2*, *TBX1*).[74] Notably, pathogenic effects on gene pathways involving *FGF8*, *LEFTY*, and *NODAL*[77–80] can contribute to left–right asymmetry disorders such as heterotaxy, a prominent CHD phenotype in patients with *SMAD2* variants. We suggest that these epigenetic and transcriptional changes provide a mechanism by which haploinsufficiency of *SMAD2* cause CHD and vascular anomalies.

SMAD2 regulates the expression of many more genes[9,10,17] than these ~200 target genes with altered expression in SMAD2-haploinsufficent iPS cells. We noticed that expression of some SMAD2/3 target genes is altered in SMAD2-haploinsufficient cells compared with other SMAD2/3 target genes. Our analyses of enriched transcription factor motifs in SMAD2-haploinsufficient cells provides insights into this observation. It appears that the affinity of SMAD2/3 for its binding sites, in open or closed chromatin, depends in large part on the nucleotide sequence of the SMAD3 binding site and the proximity of other transcription factors, such as FOXH1, NANOG and ETS, reflected by the association of binding sites in close proximity to the SMAD2/3 binding site. Our analysis of differential ATAC peaks in *SMAD2*-haploinsufficient cells demonstrates no enrichment for specific SMAD2/3 binding sites (ie, SMAD2/3 binds the same DNA sequences regardless of the amount of SMAD2/3 protein); however, there is enrichment of nearby transcription factor binding sites. That is, SMAD2/3 binding to its target sequence is determined in large part by the other transcription factors bound near the target sequence. For example, with 50% reduction in the amount of SMAD2, there are ~4 fold more differential SMAD-bound open chromatin peaks with TEAD3/4 binding in *SMAD2*-haploinsufficient cells than in SMAD-bound open chromatin peaks from WT cells (12.19% versus 3.11%; Figure 3). Whether the binding specificity of all transcription factors is controlled primarily by the proximity of other transcriptional regulators, as observed for SMAD2, or by the nucleotide sequence of the specific transcription factor binding site remains an unresolved question.

Exome analysis of samples from CHD probands also identified 24 heterozygous *SMAD2* missense variants of uncertain significance, including W274C (MH2 domain) and R114C (MH1 domain) residues. Based on the epigenetic and transcriptional changes observed in these missense lines, we deduced that both are likely pathogenic. *SMAD2*$^{+/W274C}$ iPSCs had similar loss of open chromatin peaks and comparable numbers of dysregulated genes as *SMAD2*$^{+/-}$ lines. Dysregulated expression of direct SMAD targets (n=128) included *COL5A2* and other CHD genes (*FGF8* and *ADAMTS10*). As pathogenic variants in *COL5A2* cause Ehlers Danlos syndrome,[74,81,82] a vascular syndrome with high frequency of arterial malformations and aneurysms previously associated in adult patients with *SMAD2* missense variants located in the MH2 domain,[44–49] we suggest that misexpression of *COL5A2* is key to these phenotypes.

In contrast, the R114C variant altered fewer chromatin peaks than other *SMAD2* mutant lines and transcription factor binding motifs were unchanged from those in WT cells. However, 52 genes were dysregulated, including 1 direct SMAD target and CHD gene, *GLI3*, that participates in hedgehog signaling, a critical determinant of for left–right axis formation and heart development.[83,84] Consistent with this observation, *SMAD2*$^{+/R114C}$ iPSCs had dysregulated expression including *FGF8*, *LEFTY*, and *NODAL*, findings that provide a possible mechanism for the shared clinical phenotypes in heterozygous null, W274C and R114C SMAD2 variants.

Conclusions

In conclusion, by defining mechanisms by which *SMAD2* variants disrupt epigenetic and transcriptional networks, we identified key dysregulated target genes that are critical for cardiac and vascular development and function. These findings further imply that continued monitoring of CHD probands with *SMAD2* variants is warranted to proactively identify and prevent deleterious outcomes associated with vascular aneurysms. The use of genetically engineered models provides insights into the pathogenesis of diseases such as CHD.

ARTICLE INFORMATION

Received November 5, 2024; accepted January 17, 2025.

Affiliations

Department of Genetics, Harvard Medical School, Boston, MA (T.W., S.U.M., G.V., W.T., M.Y.J., J.G., D.D., L.K.W., Z.K., J.H., S.R.D., C.S., J.G.S.); Division of Newborn Medicine, Boston Children's Hospital, Boston, MA (S.U.M.); Cardurion Pharmaceuticals, Inc., Burlington, MA (J.H.); Mindich Child Health and Development Institute and the Department of Pediatrics and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY (B.D.G.); Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA (W.K.C.); Gladstone Institutes, San Francisco, CA (B.G.B.); Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA (B.G.B.); Department of Pediatrics, Cardiovascular Research Institute, Institute for Human Genetics, Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA (B.G.B.); Department of Genetics and Pediatrics, Yale University School of Medicine, New Haven, CT (M.B.); Division of Pediatric Cardiology, University of Utah and School of Medicine, Salt Lake City, UT (M.T.); Department of Medicine, Brigham and Women's Hospital, Boston, MA (C.S.); and Howard Hughes Medical Institute, Harvard Medical School, Boston, MA (C.S.).

Acknowledgments

The authors would like to thank Paula Montero Llopis of the MicRoN (Microscopy Resources On the North Quad) core for her support and assistance, specifically for collecting images in Figure 2. We appreciate the study participants and their families, without whom this work would not have been possible. We gratefully acknowledge the National Heart, Lung, and Blood Institute Pediatrics Cardiac Genomics Consortium and Cardiovascular Development Consortium investigators (Bruce D. Gelb, Wendy K. Chung, Martina Brueckner, Martin Tristani-Firouzi, Benoit G. Bruneau, Christine Seidman, J. G. Seidman) for their support and expertise in cardiovascular development.

Sources of Funding

Funding support for this study was provided by grants to the Pediatrics Cardiac Genomics Consortium and Cardiovascular Development Consortium by the US National Heart Lung and Blood Institute (UM1 HL098123 [Bruce D. Gelb], UM1HL098179 [Benoit G. Bruneau], R01 HL151257, 1R01HL162356 [Christine Seidman], 1UM1HL098166 [J. G. Seidman]),National Institutes of Health Ruth L. Kirschstein National Research Service Award 2T32 HL 7208-46 A1 (Tarsha Ward), Stanley J. Sarnoff Cardiovascular Research Foundation (Warren Tai), Howard Hughes Medical Institute Medical Research Fellowship (Min Young Jang). Funding was supported by the Howard Hughes Medical Institute (Christine Seidman), and Foundation Leducq 16 CVD 03 (J. G. Seidman).

Disclosures

None.

Supplemental Material

Data S1
Tables S1–S7
Figures S1–S6
Videos S1–S5

REFERENCES

1. van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, Roos-Hesselink JW. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol*. 2011;58:2241–2247. doi: 10.1016/j.jacc.2011.08.025
2. Hoffman JIE, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol*. 2002;39:1890–1900. doi: 10.1016/S0735-1097(02)01886-7
3. Liu Y, Chen S, Zühlke L, Black GC, Choy M-K, Li N, Keavney BD. Global birth prevalence of congenital heart defects 1970–2017: updated systematic review and meta-analysis of 260 studies. *Int J Epidemiol*. 2019;48:455–463. doi: 10.1093/ije/dyz009
4. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, DePalma SR, McKean D, Wakimoto H, Gorham J, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*. 2015;350:1262–1266. doi: 10.1126/science.aac9396
5. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, et al. Contribution of rare inherited and de novo variants in 2871 congenital heart disease probands. *Nat Genet*. 2017;49:1593–1601. doi: 10.1038/ng.3970
6. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498:220–223. doi: 10.1038/nature12141
7. Morton SU, Quiat D, Seidman JG, Seidman CE. Genomic frontiers in congenital heart disease. *Nat Rev Cardiol*. 2022;19:26–42. doi: 10.1038/s41569-021-00587-4

8. Ward T, Tai W, Morton S, Impens F, Van Damme P, Van Haver D, Timmerman E, Venturini G, Zhang K, Jang MY, et al. Mechanisms of congenital heart disease caused by NAA15 haploinsufficiency. *Circ Res*. 2021;128:1156–1169. doi: 10.1161/CIRCRESAHA.120.316966
9. Massagué J, Seoane J, Wotton D. Smad transcription factors. *Genes Dev*. 2005;19:2783–2810. doi: 10.1101/gad.1350705
10. Yang J, Jiang W. The role of SMAD2/3 in human embryonic stem cells. *Front Cell Dev Biol*. 2020;8:653. doi: 10.3389/fcell.2020.00653
11. Moustakas A, Heldin C-H. The regulation of TGFbeta signal transduction. *Development*. 2009;136:3699–3714. doi: 10.1242/dev.030338
12. Euler-Taimor G, Heger J. The complex pattern of SMAD signaling in the cardiovascular system. *Cardiovasc Res*. 2006;69:15–25. doi: 10.1016/j.cardiores.2005.07.007
13. Zou M-L, Chen Z-H, Teng Y-Y, Liu S-Y, Jia Y, Zhang K-W, Sun ZL, Wu JJ, Yuan ZD, Feng Y, et al. The Smad dependent TGF-β and BMP signaling pathway in bone remodeling and therapies. *Front Mol Biosci*. 2021;8:593310. doi: 10.3389/fmolb.2021.593310
14. Nickel J, Mueller TD. Specification of BMP signaling. *Cells*. 2019;8:1579. doi: 10.3390/cells8121579
15. Shen MM. Nodal signaling: developmental roles and regulation. *Development*. 2007;134:1023–1034. doi: 10.1242/dev.000166
16. Barnes RM, Black BL. Nodal signaling and congenital heart defects. In: Nakanishi T, Markwald RR, Baldwin HS, Keller BB, Srivastava D, Yamagishi H, eds *Etiology and Morphogenesis of Congenital Heart Diseases*. Springer Japan; 2016:183–192. doi: 10.1007/978-4-431-54628-3_24
17. Massague J. TGFbeta signalling in context. *Nat Rev Mol Cell Biol*. 2012;13:616–630. doi: 10.1038/nrm3434
18. Nakao A, Röijer E, Imamura T, Souchelnytskyi S, Stenman G, Heldin C-H, Dijke P. Identification of Smad2, a human mad-related protein in the transforming growth factor β signaling pathway. *J Biol Chem*. 1997;272:2896–2900. doi: 10.1074/jbc.272.5.2896
19. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature*. 2003;425:577–584. doi: 10.1038/nature02006
20. Takenoshita S, Mogi A, Nagashima M, Yang K, Yagi K, Hanyu A, Nagamachi Y, Miyazono K, Hagiwara K. Characterization of the MADH2/Smad2 gene, a human mad homolog responsible for the transforming growth factor-beta and activin signal transduction pathway. *Genomics*. 1998;48:1–11. doi: 10.1006/geno.1997.5149
21. Lo RS, Chen YG, Shi Y, Pavletich NP, Massagué J. The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-beta receptors. *EMBO J*. 1998;17:996–1005. doi: 10.1093/emboj/17.4.996
22. Shi Y, Wang YF, Jayaraman L, Yang H, Massagué J, Pavletich NP. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell*. 1998;94:585–594. doi: 10.1016/S0092-8674(00)81600-1
23. Wu JW, Hu M, Chai J, Seoane J, Huse M, Li C, Rigotti DJ, Kyin S, Muir TW, Fairman R, et al. Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. *Mol Cell*. 2001;8:1277–1289. doi: 10.1016/S1097-2765(01)00421-X
24. Kim SW, Yoon S-J, Chuong E, Oyolu C, Wills AE, Gupta R, Baker J. Chromatin and transcriptional signatures for nodal signaling during endoderm formation in hESCs. *Dev Biol*. 2011;357:492–504. doi: 10.1016/j.ydbio.2011.06.009
25. Hill CS. Transcriptional control by the SMADs. *Cold Spring Harb Perspect Biol*. 2016;8:1–17. doi: 10.1101/cshperspect.a022079
26. Coda DM, Gaarenstroom T, East P, Patel H, Miller DSJ, Lobley A, Matthews N, Stewart A, Hill CS. Distinct modes of SMAD2 chromatin binding and remodeling shape the transcriptional response to NODAL/Activin signaling. *eLife*. 2017;6:1–31. doi: 10.7554/eLife.22474
27. Coda DM, Patel H, Gori I, Gaarenstroom TE, Song O-R, Howell M, Hill CS. A network of transcription factors governs the dynamics of NODAL/Activin transcriptional responses. *J Cell Sci*. 2022;135:1–17. doi: 10.1242/jcs.259972
28. Massagué J, Xi Q. TGF-β control of stem cell differentiation genes. *FEBS Lett*. 2012;586:1953–1958. doi: 10.1016/j.febslet.2012.03.023
29. Gaarenstroom T, Hill CS. TGF-beta signaling to chromatin: how Smads regulate transcription during self-renewal and differentiation. *Semin Cell Dev Biol*. 2014;32:107–118. doi: 10.1016/j.semcdb.2014.01.009
30. Beyer TA, Weiss A, Khomchuk Y, Huang K, Ogunjimi AA, Varelas X, Wrana JL. Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells. *Cell Rep*. 2013;5:1611–1624. doi: 10.1016/j.celrep.2013.11.021
31. Labbé E, Silvestri C, Hoodless PA, Wrana JL, Attisano L. Smad2 and Smad3 positively and negatively regulate TGF beta-dependent transcription through the forkhead DNA-binding protein FAST2. *Mol Cell*. 1998;2:109–120. doi: 10.1016/S1097-2765(00)80119-7
32. Mullen AC, Orlando DA, Newman JJ, Loven J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell*. 2011;147:565–576. doi: 10.1016/j.cell.2011.08.050
33. Xu R-H, Sampsell-Barron TL, Gu F, Root S, Peck RM, Pan G, Yu J, Antosiewicz-Bourget J, Tian S, Stewart R, et al. NANOG is a direct target of TGFβ/Activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell*. 2008;3:196–206. doi: 10.1016/j.stem.2008.07.001
34. Brown S, Teo A, Pauklin S, Hannan N, Cho CH-H, Lim B, Vardy L, Dunn NR, Trotter M, Pedersen R, et al. Activin/nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells*. 2011;29:1176–1185. doi: 10.1002/stem.666
35. Liberati NT, Datto MB, Frederick JP, Shen X, Wong C, Rougier-Chapman EM, Wang XF. Smads bind directly to the Jun family of AP-1 transcription factors. *Proc Natl Acad Sci USA*. 1999;96:4844–4849. doi: 10.1073/pnas.96.9.4844
36. Moustakas A, Pardali K, Gaal A, Heldin CH. Mechanisms of TGF-beta signaling in regulation of cell growth and differentiation. *Immunol Lett*. 2002;82:85–91. doi: 10.1016/s0165-2478(02)00023-8
37. Ding A, Bian Y-Y, Zhang Z-H. SP1/TGF-β1/SMAD2 pathway is involved in angiogenesis during osteogenesis. *Mol Med Rep*. 2020;21:1581–1589. doi: 10.3892/mmr.2020.10965
38. Pardali K, Kurisaki A, Morén A, ten Dijke P, Kardassis D, Moustakas A. Role of Smad proteins and transcription factor Sp1 in p21Waf1/Cip1 regulation by transforming growth factor-β *. *J Biol Chem*. 2000;275:256. doi: 10.1074/jbc.M909467199
39. Sundqvist A, Zieba A, Vasilaki E, Herrera Hidalgo C, Söderberg O, Koinuma D, Miyazono K, Heldin CH, Landegren U, ten Dijke P, et al. Specific interactions between Smad proteins and AP-1 components determine TGFβ-induced breast cancer cell invasion. *Oncogene*. 2013;32:3606–3615. doi: 10.1038/onc.2012.370
40. Koinuma D, Tsutsumi S, Kamimura N, Taniguchi H, Miyazawa K, Sunamura M, Imamura T, Miyazono K, Aburatani H. Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. *Mol Cell Biol*. 2009;29:172–186. doi: 10.1128/MCB.01038-08
41. Katoh M, Katoh M. Integrative genomic analyses of ZEB2: transcriptional regulation of ZEB2 based on SMADs, ETS1, HIF1alpha, POU/OCT, and NF-kappaB. *Int J Oncol*. 2009;34:1737–1742. doi: 10.3892/ijo_00000304
42. Zhang Z, Zhang F, Zhang M, Xue H, Fan L, Weng Y. The role of SMAD signaling in hypertrophic obstructive cardiomyopathy: an immunohistopathological study in pediatric and adult patients. *Sci Rep*. 2023;13:3706. doi: 10.1038/s41598-023-30776-9
43. Zhong H, Zhang R, Li G, Huang P, Zhang Y, Zhu J, Kuang J, Hutchins AP, Qin D, Zhu P, et al. c-JUN is a barrier in hESC to cardiomyocyte transition. *Life Sci Alliance*. 2023;6(11):e202302121. doi: 10.26508/lsa.202302121
44. Granadillo JL, Chung WK, Hecht L, Corsten-Janssen N, Wegner D, Nij Bijvank SWA, Toler TL, Pineda-Alvarez DE, Douglas G, Murphy JJ, et al. Variable cardiovascular phenotypes associated with SMAD2 pathogenic variants. *Hum Mutat*. 2018;39:1875–1884. doi: 10.1002/humu.23627
45. Micha D, Guo D-C, Hilhorst-Hofstee Y, van Kooten F, Atmaja D, Overwater E, Cayami FK, Regalado ES, van Uffelen R, Venselaar H, et al. SMAD2 mutations are associated with arterial aneurysms and dissections. *Hum Mutat*. 2015;36:1145–1149. doi: 10.1002/humu.22854
46. Cannaerts E, Kempers M, Maugeri A, Marcelis C, Gardeitchik T, Richer J, Micha D, Beauchesne L, Timmermans J, Vermeersch P, et al. Novel pathogenic SMAD2 variants in five families with arterial aneurysm and dissection: further delineation of the phenotype. *J Med Genet*. 2019;56:220–227. doi: 10.1136/jmedgenet-2018-105304
47. Zhang W, Zeng Q, Xu Y, Ying H, Zhou W, Cao Q, Zhou W. Exome sequencing identified a novel SMAD2 mutation in a Chinese family with early onset aortic aneurysms. *Clin Chim Acta*. 2017;468:211–214. doi: 10.1016/j.cca.2017.03.007
48. Vandeloo B, Azzano A, Schoors D, Verstraeten A, Van Laer L, Loeys B, Vermeersch P. Spontaneous coronary artery dissection in a man

with a novel missense mutation in SMAD2 treated by optical coherence tomography-guided percutaneous coronary intervention. *JACC Cardiovasc Interv*. 2019;12:e45–e47. doi: 10.1016/j.jcin.2018.09.007
49. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980–D985. doi: 10.1093/nar/gkt1113
50. Gelb B, Brueckner M, Chung W, Goldmuntz E, Kaltman J, Kaski JP, Kim R, Kline J, Mercer-Rosa L. The congenital heart disease genetic network study. *Circ Res*. 2013;112:698–706. doi: 10.1161/circresaha.111.300297
51. Ohye RG, Sleeper LA, Mahony L, Newburger JW, Pearson GD, Lu M, Goldberg CS, Tabbutt S, Frommelt PC, Ghanayem NS, et al. Comparison of shunt types in the Norwood procedure for single-ventricle lesions. *N Engl J Med*. 2010;362:1980–1992. doi: 10.1056/NEJMoa0912461
52. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393
53. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303. doi: 10.1101/gr.107524.110
54. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11101–11133. doi: 10.1002/0471250953.bi1110s43
55. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92. doi: 10.4161/fly.19695
56. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164. doi: 10.1093/nar/gkq603
57. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. The mutational constraint spectrum quantified from variation in 14 1456 humans. *Nature*. 2020;581:434–443. doi: 10.1038/s41586-020-2308-7
58. Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, Kamp TJ, Palecek SP. Cozzarelli prize winner: robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci USA*. 2012;109:E1848–E1857. doi: 10.1073/pnas.1200250109
59. Lian X, Zhang J, Azarin SM, Zhu K, Hazeltine LB, Bao X, Hsiao C, Kamp TJ, Palecek SP. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. *Nat Protoc*. 2013;8:162–175. doi: 10.1038/nprot.2012.150
60. Toepfer CN, Sharma A, Cicconet M, Garfinkel AC, Mucke M, Neyazi M, Willcox JA, Agarwal R, Schmid M, Rao J, et al. SarcTrack. *Circ Res*. 2019;124:1172–1183. doi: 10.1161/CIRCRESAHA.118.314505
61. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1–21.29.9. doi: 10.1002/0471142727.mb2129s109
62. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16:284–287. doi: 10.1089/omi.2011.0118
63. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381:eadg7492.
64. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47:D886–D894. doi: 10.1093/nar/gky1016
65. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24:2125–2137. doi: 10.1093/hmg/ddu733
66. Sakaki-Yumoto M, Liu J, Ramalho-Santos M, Yoshida N, Derynck R. Smad2 is essential for maintenance of the human and mouse primed pluripotent stem cell state. *J Biol Chem*. 2013;288:560. doi: 10.1074/jbc.M112.446591
67. Kitamura R, Takahashi T, Nakajima N, Isodono K, Asada S, Ueno H, Ueyama T, Yoshikawa T, Matsubara H, Oh H. Stage-specific role of endogenous Smad2 activation in Cardiomyogenesis of embryonic stem cells. *Circ Res*. 2007;101:78–87. doi: 10.1161/circresaha.106.147264
68. Dunn NR, Vincent SD, Oxburgh L, Robertson EJ, Bikoff EK. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development*. 2004;131:1717–1728. doi: 10.1242/dev.01072
69. Faial T, Bernardo AS, Mendjan S, Diamanti E, Ortmann D, Gentsch GE, Mascetti VL, Trotter MWB, Smith JC, Pedersen RA. Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development*. 2015;142:2121–2135. doi: 10.1242/dev.117838
70. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, Kathiriya IS, Hinson JT, Homsy J, Gray J, Pu W, et al. Single-cell resolution of temporal gene expression during heart development. *Dev Cell*. 2016;39:480–490. doi: 10.1016/j.devcel.2016.10.001
71. Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, et al. The gene ontology knowledgebase in 2023. *Genetics*. 2023224:1–14. doi: 10.1093/genetics/iyad031
72. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25–29. doi: 10.1038/75556
73. Besser D. Expression of nodal, lefty-a, and lefty-B in undifferentiated human embryonic stem cells requires activation of Smad2/3. *J Biol Chem*. 2004;279:84. doi: 10.1074/jbc.M404979200
74. McKusick VA. *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*. Johns Hopkins University Press; 1998.
75. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–589. doi: 10.1016/j.molcel.2010.05.004
76. Li Y, Zhang Y, Shi H, Liu X, Li Z, Zhang J, Wang X, Wang W, Tong X. CRTC2 activates the epithelial-mesenchymal transition of diabetic kidney disease through the CREB-Smad2/3 pathway. *Mol Med*. 2023;29:146. doi: 10.1186/s10020-023-00744-0
77. Hamada H, Meno C, Watanabe D, Saijoh Y. Establishment of vertebrate left–right asymmetry. *Nat Rev Genet*. 2002;3:103–113. doi: 10.1038/nrg732
78. García-Castro MI, Vielmetter E, Bronner-Fraser M. N-cadherin, a cell adhesion molecule involved in establishment of embryonic left–right asymmetry. *Science*. 2000;288:1047–1051.
79. Catana A, Apostu AP. The determination factors of left–right asymmetry disorders-a short review. *Clujul Med*. 2017;90:139–146. doi: 10.15386/cjmed-701
80. Albertson RC, Yelick PC. Roles for fgf8 signaling in left–right patterning of the visceral organs and craniofacial skeleton. *Dev Biol*. 2005;283:310–321. doi: 10.1016/j.ydbio.2005.04.025
81. Malfait F, Wenstrup RJ, De Paepe A. Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. *Genet Med*. 2010;12:597–605. doi: 10.1097/GIM.0b013e3181eed412
82. Richards AJ, Martin S, Nicholls AC, Harrison JB, Pope FM, Burrows NP. A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type II. *J Med Genet*. 1998;35:846–848. doi: 10.1136/jmg.35.10.846
83. Bijlsma MF, Peppelenbosch MP, Spek CA. Hedgehog morphogen in cardiovascular disease. *Circulation*. 2006;114:1985–1991. doi: 10.1161/CIRCULATIONAHA.106.619213
84. Zhang XM, Ramalho-Santos M, McMahon AP. Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node. *Cell*. 2001;105:781–792.